

SUMMARIES

(without alignments)

6189.297 Million cel

.....GKIARVRRARRRR

residues

Parameters: 3230813

022003 16431E 1263/32

ATCH=0.1 - LOOPCL=0 - L

-THR_MIN=0 -ALIGN=15
-MAYT.EN=2000000000

```
0 -WATT -FONGIOG -DE
2003_164315_1262 -NCP
```

```

APEXT=0.5 -FGAPOP=6 -

```

Result No.	Score	Query Match	Length	DB	ID	Description
1	121.5	3.6	3775	11	BC028165	BC028165 Homo sapi
2	116	3.4	1535	13	BM552027	BM552027 AGENCOURT
3	112	3.3	900	12	BF167575	BF167575 601774283
4	111.5	3.3	1069	14	BM913440	BM913440 AGENCOURT
5	111.5	3.3	1269	14	BQ712831	BQ712831 AGENCOURT
6	109.5	3.2	2128	13	BI0822474	BI0822474 602877815
7	109	3.2	3572	11	BC010575	BC010575 Homo sapi
8	108.5	3.2	915	13	BI413164	BI413164 60290033
9	108.5	3.2	1776	14	BM926649	BM926649 AGENCOURT
10	108	3.2	861	12	BG168891	BG168891 602320281
11	107.5	3.2	869	9	AL578091	AL578091 AL578091
12	106.5	3.2	686	17	AG139016	AG139016 Pan trogl
13	106.5	3.2	820	17	BI650823	BI650823 603297718
14	106	3.1	959	17	CNS0303N	AL250484 Tetraodon
15	106	3.1	983	14	BQ649851	BQ649851 AGENCOURT
16	106	3.1	1060	13	BM549598	BM549598 AGENCOURT
17	106	3.1	1061	17	CNS03379	AL252574 Tetraodon
18	105.5	3.1	635	9	A1541596	A1541596 SD02772.5
19	105.5	3.1	783	9	BI652061	BI652061 603298842
20	105.5	3.1	1090	13	BM479488	BM479488 AGENCOURT
21	105.5	3.1	1294	13	BI409855	BI409855 602961973
22	105	3.1	1695	13	BJ244281	BJ244281 BJ244281
23	105	3.1	1087	12	BF973817	BF973817 602241914
24	105	3.1	1485	14	BM907789	BM907789 AGENCOURT
25	104.5	3.1	968	12	BE736039	BE736039 601305847
26	104.5	3.1	962	14	BQ935611	BQ935611 AGENCOURT
27	104	3.1	891	14	BQ670497	BQ670497 AGENCOURT
28	104	3.1	973	12	BF981727	BF981727 602306079
29	104	3.1	978	14	BQ714511	BQ714511 AGENCOURT
30	104	3.1	2391	17	BH771003	BH771003 LTMG-a972
31	103.5	3.1	679	13	BM406277	BM406277 EST580604
32	103.5	3.1	830	14	BM995604	BM995604 2_A12-T3
33	103.5	3.1	963	14	BQ899717	BQ899717 AGENCOURT
34	103.5	3.1	1067	12	BG168781	BG168781 602320073
35	103.5	3.1	1092	14	BQ676763	BQ676763 AGENCOURT
36	103.5	3.1	1858	13	BM557460	BM557460 AGENCOURT
37	103	3.1	937	13	BM461960	BM461960 AGENCOURT
38	103	3.1	1031	17	CNS0519X	AL439270 Tetraodon
39	102.5	3.0	663	14	BQ917203	BQ917203 QHB20J22.
40	102.5	3.0	750	13	BI768562	BI768562 603057176
41	102.5	3.0	942	14	BQ930180	BQ930180 AGENCOURT
42	102.5	3.0	1060	13	BM424198	BM424198 AGENCOURT
43	102.5	3.0	1287	13	BM564626	BM564626 AGENCOURT
44	102	3.0	1259	17	CNS033PT	AL251531 Tetraodon
45	102	3.0	1259	14	BQ960839	BQ960839 AGENCOURT

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION
BC028165	3775 bp	mRNA
BC028165	Homo sapiens, protocadherin 17, clone IMAGE:5245587, mRNA.	linear HTC 01-MAY-2002

VERSION BC028165.1 GI:20380902

SOURCE	Homo sapiens.
1. The Bible	
2. The Quran	
3. The Vedas	
4. The Upanishads	
5. The Bhagavad Gita	
6. The Ramayana	
7. The Mahabharata	
8. The Puranas	
9. The Arthashastra	
10. The Kautilya's Arthashastra	
11. The Manusmriti	
12. The Dharma Smritis	
13. The Smritis	
14. The Smritis	
15. The Smritis	
16. The Smritis	
17. The Smritis	
18. The Smritis	
19. The Smritis	
20. The Smritis	
21. The Smritis	
22. The Smritis	
23. The Smritis	
24. The Smritis	
25. The Smritis	
26. The Smritis	
27. The Smritis	
28. The Smritis	
29. The Smritis	
30. The Smritis	
31. The Smritis	
32. The Smritis	
33. The Smritis	
34. The Smritis	
35. The Smritis	
36. The Smritis	
37. The Smritis	
38. The Smritis	
39. The Smritis	
40. The Smritis	
41. The Smritis	
42. The Smritis	
43. The Smritis	
44. The Smritis	
45. The Smritis	
46. The Smritis	
47. The Smritis	
48. The Smritis	
49. The Smritis	
50. The Smritis	
51. The Smritis	
52. The Smritis	
53. The Smritis	
54. The Smritis	
55. The Smritis	
56. The Smritis	
57. The Smritis	
58. The Smritis	
59. The Smritis	
60. The Smritis	
61. The Smritis	
62. The Smritis	
63. The Smritis	
64. The Smritis	
65. The Smritis	
66. The Smritis	
67. The Smritis	
68. The Smritis	
69. The Smritis	
70. The Smritis	
71. The Smritis	
72. The Smritis	
73. The Smritis	
74. The Smritis	
75. The Smritis	
76. The Smritis	
77. The Smritis	
78. The Smritis	
79. The Smritis	
80. The Smritis	
81. The Smritis	
82. The Smritis	
83. The Smritis	
84. The Smritis	
85. The Smritis	
86. The Smritis	
87. The Smritis	
88. The Smritis	
89. The Smritis	
90. The Smritis	
91. The Smritis	
92. The Smritis	
93. The Smritis	
94. The Smritis	
95. The Smritis	
96. The Smritis	
97. The Smritis	
98. The Smritis	
99. The Smritis	
100. The Smritis	

Eukaryota; Me

REFERENCE 1 (bases 1 to 3775)

TITLE	Direct Submission

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 REMARK
 COMMENT
 Contact: MGC help desk
 Email: gcgabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Bikesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Districh, N.L., Granite, S., Guan, X., Gupta, J., Haight, P., Hansen, N., Ho, S.-L., Karlins, E., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stanirlop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 62 Row: n Column: 5
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gt: 14589926
 This clone has the following problem: frame shifted.

FEATURES

SOURCE

Location/Qualifiers
 1. 3775
 /organism="Homo sapiens"
 /db_xref="LocusID:27253"
 /db_xref="taxon:9606"
 /clone="IMAGE:5245587"
 /tissue_type="Brain, fetal, whole pooled"
 /clone_id="NH_MGC_121"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

BASE COUNT 916 a 1117 c 1031 g 711 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.648 Length: 3775
 Score: 121.50 Matches: 158
 Percent Similarity: 35.20% Conservative: 93
 Best Local Similarity: 22.16% Mismatches: 248
 Query Match: 3.60% Indels: 216
 Gaps: 38

US-09-677-653A-50 (1-647) x BC028165 (1-3775)

QY 12 HtAsnAaGArGgLyThrArGAsnValArGValSerAlaAsn-ThrValThValAsnG 31
 Db 318 CACAAATGCGCAAGTGCACGCTCTCGAGGTGTCGCCAACGACA----- 363
 QY 31 YArGArGAsnGln-----ArGArGArGThrGlnArGValSerProProAs 47
 Db 364 -AGGAGATCTGCATGATCAAGTAGAGATCCAGGACATCAACGACGCGCTCTTC 422
 QY 47 pAsnPherAlaAlaAlaGlnAspLeuAlaGlnSerLeuAspAlaAsnThrValThPh 67
 Db 423 TCCTCGACACGATCGAA-ATGGACATCTCGGAGAACGCTCTCGGCGACC---CGCT 478
 QY 67 eProAlaAsnIleSerSerMetProGluPhArGAsnTrpAlaValysGlyIleAspLe 87
 Db 479 CCCCCTCACCGACGACATGACCC-----GACGC 508
 QY 87 uAspSerAspSerIleGlyTrpTyPhelAspTyIleuAspProAlaGlyAlaThGlu- 106
 Db 509 CGGGAGAAATGGCTCCGACCTACCTCTCAACGCGCAGCATCAGCGCTCTTGGA 568
 QY 107 -SerAlaArGAlaValGlyGlyTrpSerLysIleProAspGlyLeuValLysPheSerVa 126

Db 569 GGACGTTAAGTCCCGCGCGACGACCAAGTCCAGAACGATCATCAGAAAGCTCT 628
 QY 126 lAspAlaGluIleArGValIleTyAsnGluGlyProValValThAspValSerVa 146
 Db 629 GGAC-----CGGACCAACAGATATCATTACATCGCTGGCTGACT----- 668
 QY 146 lProLeuAspGlyArGValnTrpSerLeuSerIlePheSerPheProMetPheArGThAl 166
 Db 669 -GCCCTGGACGGTGGCGAG-----CCTCCAGCTTCGCGACCGT 706
 QY 166 aTyArValAlaValAlaAsnValGluAsnLysGluMetSer-----LeuAspValValAs 184
 Db 707 ACAGATCAACGTAAGGATGATGATCAACCAACAGACCCGCTTCGAGCGCCCATC 766
 QY 184 nAspLeuIleGlyTrpPheAsnAsnLeuAlaAspTrpArGValValAspSerGlu 204
 Db 767 CTACTGGTGGAGATCCCGGACAGACGCTCCGTGGTACAGTGGTCATGAT----- 818
 QY 204 nTrpIleAsnPherTrasn-----AspTrnTrpTyf----- 214
 Db 819 ---CTGAACGCCACCGACCGCATGAGTCCCATATGATGATGATGATGATGATGAT 874
 QY 215 ---TyArValArGValArGValLeuArGProThrTyArSpValProAspProThrGlu 233
 Db 875 CAGCTACGCTGCTGACCGCGT---CGGAGCTCTTTCATC---GACCCCAAGACCG 928
 QY 233 yLeuValArGThrValSer-----AspTyf----- 241
 Db 929 CCTAATCGGTGAAGGCAATCTGACTATGAGAAACCGGATCTGACATTGACGT 988
 QY 242 -----ArGLeuThrTyfLysAl 247
 Db 989 GCAGGCCGAGACCTGGGCGCTTAACCTATCCACCCATCCAAAGTACACGCTAAC 1048
 QY 247 aIleThrCysGluAlaAsnMetProThrLeuValAspGlnGlyTrpIleGlyGlu 267
 Db 1049 CATGCACGCAACGACAAATGCGCGTCCATC-----GGTTC----- 1085
 QY 267 nTyArAlaLeuThrProThrSerLeuProGlnTrpArSpValSerGluAlaLeuH 287
 Db 1086 -----GTCTCGTCCGCGGAGGCGCGCTGACGAGGCC----- 1118
 QY 287 sThrLeuThrPheAlaArGProSerSerAlaAlaLeuAlaPheValTrp----- 304
 Db 1119 -----GGCCCTCCGCGACCGTATCCCTGTGGCGGCGACAGACCGGGA 1165
 QY 305 -AlaGly-----LeuProGlnGlyGlyThrAlaProAl 315
 Db 1166 CTCTGCAAGAACGACAGCTTCAGTCTCGGTCCTAGCGGAGGAGGCGCGCGCG 1225
 QY 315 aGlyThrProAlaTrpGluGlnAlaSerSerGlyLyIleThrTrpArGHisAsnG 335
 Db 1226 CGG-----G-----G-----G-----G-----G----- 1231
 QY 335 yThrThrPheProAlaGlySerValSerTyArValLeuProGluGlyPheAlaLeuGluAr 355
 Db 1232 CCTGGCGCGGCGCGGGGTTCGTCCTTCATCAAGCTTAGAGAGAACTAC----- 1280
 QY 355 gTyArSpProAsnAspGlySerTrpThrAspPheAlaSerAlaGlyAspTrnValThPh 375
 Db 1281 -----GACACTTCTACACG-----GTGGTGACATGA 1306
 QY 375 eArGlnVal-----AlaValAspGluValValThrAsnAsnProAlaGlyG 392
 Db 1307 CCGCGCGCTGACCGCGACACAGACAGAGATACCACTGATCTGGCGCGGACGG 1366
 QY 392 yGlySerAlaPro-----ThrPheThrValArGVal----- 402
 Db 1367 GGGCTCTCTCCCTCCCACTCCACCAAGTCTGCGGATCAAGATTCAGACGAGAA 1426
 QY 403 ---ProPro-SerAsnAlaTyTrhAsnThrValPheArGAsnThrLeuGluThra 421

Db 1427 CAACCCGCGCTGACCAAGGCTTACGCTTACAGTGCACAGAACAACTCC 1486
 Qy 421 rGProSerSerArgATgLeuGluLeuProMetProProAlaSerPheGluGlnThVala 441
 Db 1487 GGAGAGTACTGCTGGCTGTG-CTCGCCAGAGATCC--GACCTGGCCAG----- 1535
 Qy 441 laasnProFroYlIleGluGlnSerLeuLeuGlnThVala 461
 Db 1536 -----AACGGCAACCGATCTACTCTATCTCTGCGCCATCGGCGAGCTGTATCT 1590
 Qy 461 lSerLeuMetArg--AsnProValPheGlnLeuThrProAlaSerSerPheGluVala 480
 Db 1591 ACACCTATGTGTGTGAATCC-----ACGAACGGGGCCATCTACGCCCTGC 1638
 Qy 480 alSerPheAsnAsnProGlyTyrGluArgThrArgAspLeuProAspTyrThrGlyIle 500
 Db 1639 GCTCTTTAAC-----TTGAGCAGACCAAGGCTTTTGAGTTCAGAGTCTTCTA 1689
 Qy 500 rGAspSer-----PheAspGlnAsnMetSerThrAlaValAlaIleAspPhe 515
 Db 1690 AGACACTGGGGGGGCGCCGCGCACTTGAGAGCAACGCCAGCTGAGGTGACAGTCTAG 1749
 Qy 515 rGSerLeuSerHisSerCysSerIleValThrIysThrTyrGlnGlyTyrGluGly---- 533
 Db 1750 ACCTGAATGACAACGCGCAGATGATGCTCTCCACAGCTGCAGAACGACACCGCGAGC 1809
 Qy 534 -----ValThrAsnValAsnThrProPheG 542
 Db 1810 TGCAGGTGCCCGCAACGCTGCGCTGCTATCTGTGAGCACTGTGCGCCCTAGACA 1869
 Qy 542 lYglNpHeAlaHisAlaGlyLeuLeuLys-----AsnGluGluIle 556
 Db 1870 GCGACTTCGCGAGAGCGGGCGCTCTACCTACAGATCTGCGAGCGGCAACGACAGCAC 1929
 Qy 556 euCysLeuAlaAspAspLeuAlaThrArgLeuThrGlyValTyrProAlaThrAspAsn- 575
 Db 1930 TGTTTAGATTCGACCCGCTCCAGCGGCGAGTCCGACGCTCTTCTGTGGAGGACG 1989
 Qy 576 -----PheAlaAlaValSerAlaPheAlaAlaAsnMetLeuSer 590
 Db 1990 TCACGCCCGTGTGAGCTGTGCTGAGTGACGACCAACGCGCAACCTTACCTCTCCG 2049
 Qy 590 exValLeuLysSerGluAlaThrSerSerIleLeuLysSerValGlyGluThAlaVala 610
 Db 2050 CAGTGCCAG-----CTCATCATCCGCTCGTGAGC-----G 2082
 Qy 610 lYAlaAlaGlnSerGlyLeuAlaLysLeuProGly 621
 Db 2083 GATCCCTTCCCGAGGGGTACCAAGGGGTGATGGC 2117
 RESULT 2
 BMS52027 1535 bp mRNA linear EST 20-FEB-2002
 LOCUS BMS52027 5' mRNA sequence.
 DEFINITION AGENCOURT_6543677 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5549464
 ACCESSION BMS52027
 VERSION BMS52027.1 GI:16789547
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1535)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LINT at:
<http://image.llnl.gov>
 Plate: LHAM12260 row: a column: 17
 High quality sequence start: 16
 High quality sequence stop: 269.
 Location/Qualifiers
 1. 1535
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5549464"
 /clone_1ib="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="PH108 (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb.
 BASE COUNT 220 a 596 c 329 g 385 t 5 others
 ORIGIN
 Alignment Scores:
 pred. No.: 0.531 Length: 1535
 Score: 116.00 Matches: 72
 Percent Similarity: 30.28% Conservative: 37
 Best Local Similarity: 20.00% Mismatches: 114
 Query Match: 3.44% Indels: 137
 DB: 13 Gaps: 15
 US-09-677-653a-50 (1-647) x BMS52027 (1-1535)
 Qy 265 gLyGlyGlnTyrAlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyr 284
 Db 45 GAGAGTGAAGACGTGCTGAGCCAGTGGGTTCGCCACCTACCCAGAGAGCTGCTGG 104
 Qy 285 AlAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeuAlaPheValTyr 304
 Db 105 GCC-----AGACCGCGGAGAGGAGGACACTGCTGCTCTGCGCTGCTC--- 149
 Qy 305 AlAGlyLeuProGlnGlyGlyThrAlaProAlaGly---ThrProAla---TrpGluGln 322
 Db 150 -----CTTCGCGAGTGAAGGGGTGACCGAGCTTCCTTCCCATCTGTTCGAGGGA 203
 Qy 323 AlAspSerGlyGly-TyrLeuThrTrpArgHisAsnGlyThrThrPheProAlaGlyse 342
 Db 204 GGGAGAGGAGGGGCTCTC----- 222
 Qy 342 rValSerTyrValLeuProGlnGlyPheAlaLeuGluAlaGlyTyrAspProAsnAspLys 362
 Db 223 -----AGCTTGAGCCAGAGCTGGGGGTG 245
 Qy 362 rTyrPheAspPheAlaSerAlaGlyAspThrValThrPhe----- 375
 Db 246 CTGGGTGGAGATGAGATTTAGGGGGCGCCCTCGTGTCTTTTTCGCCCGGGCCAGC 305
 Qy 376 -ArgGlnValAlaValaAspGluValValThrAsnAsnProAlaGlyGlySerAl 395
 Db 306 CGGCAACCGCCCAACCGCACCTCTCCCGCAATTCGCGCGGGGGGGGGCGG 365
 Qy 395 apProThrPheThrValArgValPro--ProSer-----AsnAlaTyrThr----- 409
 Db 366 TCCCTCCCTTCCCTTCCATTCCTCATTCGCAATGTGACTCGCGCTTCACATTAGA 425
 Qy 409 ----- 409
 Db 426 CGCCATGACCCCTTGCTGTGCAACCTCACCGCAATTCCTCTCCCTCTACCTCGA 485
 Qy 410 -----AsnThrValPheArgAsnThrLeuGluThrArgProSerSerArgArgLe 427
 Db 486 TTTTTCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 545
 Qy 427 uGluLeuPro----- 430
 Db 546 CTCTGCCCTTACAGTCCGCTTCTGTCTCTTATATGCCCGCCCGCGGTGAGAGATC 605

```

Oy 431 -----MetProAlaAspPheGlyGlnThrValAlaAsnAs 443
Db 606 GGAATGTTTGNMGCGTTTNCGTTCGCCGGGACAGGGCCGACATCAGAT 665
Oy 443 nProAlaSerPheGlyAlaValSerPheAsnAsnProGlyGluArgThrArgAs 492
Db 666 TCCCGCGGACAGACATCAACACTCCCGCGGACTCTTACAGTACGATCCGCCGCC 725
Oy 456 -----GlyCysTyrLeuValHisSerLysMetArgAsnProValPheGlnLeu 472
Db 726 ACACATATACCCCGCTATCTCTCTATCTCTATCTCTATCTCTCTCTCTCTCTCT 771
Oy 472 rProAlaSerPheGlyAlaValSerPheAsnAsnProGlyGluArgThrArgAs 492
Db 772 -CCAAGCGCATCTCTCATGATGATTTGTATATCCGCAATCGGC---GACGATATGGCCA 827
Oy 492 pleu-----ProAs 495
Db 828 GCTAGACACACAGCATTTACTTCCATCCACACTCTCATATTCCTGTGTGCTCCCTC 887
Oy 495 pTyrThrGlyIleArgAspSerPheAspGlnAsnMetSer----- 508
Db 888 TTACTACAGACAGTACGCCCTTATCTCTGATCGCTCGCCCTCGCAGACAGCTCC 947
Oy 509 ---ThrAlaValAlaHisPheArgSerLeuSerHisSerCysSerIleValThrLys 526
Db 948 TGTCTGTCCTCAGTCTCGCTACACTCTACTCGCTCTCTCTCTCTCTCTCTCTCT 1005

```

RESULT 3

```

Bf167575/c 900 bp mRNA linear EST 30-OCT-2000
LOCUS 601774283p1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3994773 5',
DEFINITION mRNA sequence.
ACCESSION Bf167575
VERSION Bf167575.1 GI:11047927
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 900)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsab@femail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LAM9213 row: f column: 22
High quality sequence stop: 700.
Location/Qualifiers
1. 900

```

FEATURES

source

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clone="IMAGE:3994773"
/clone_1lb="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: Lung; Vector: pCMV-SPORT6; Site_1: Salt.
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

```

BASE COUNT

222 a 235 c 284 g 159 t

Alignment Scores:

```

Pred. No.: 0.565 Length: 900
Score: 112.00 Matches: 66
Percent Similarity: 33.33% Conserved: 25
Best Local Similarity: 24.18% Mismatches: 84
Query Match: 3.32% Indels: 98
DB: 12 Gaps: 14

```

US-09-677-653a-50 (1-647) x Bf167575 (1-900)

```

Oy 225 TyrAspValProAspProThrGluLeuValArgThrValSerAspTyrArgLeuThr 244
Db 814 TACTGTGTCGCCGATCCCGTT----- 794
Oy 245 TyrLysAlaIleThrCysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrp 263
Db 793 ---GTTGCCCCGGCTTTTATGCGCAGGCGCCCTGACGTGTGATTAAGGCTTTTCGGA 737
Oy 264 ---IleGlyGlyGlnTyrAlaLeu-----ThrPro----- 272
Db 736 GCCATTGGTGAAGTGGGGGGGCTCTCGGATTAACCTGACTTGCACCGCATTTGATGG 677
Oy 273 -----ThrSerLeuProGlnTyrAspValSerGluAlaTyr 284
Db 676 AAGAGCCAGCAGTGCAGTTTGAAGCTCTTACCCAGGCGCCGCCACACAGACT--- 620
Oy 285 AlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaLeuAlaPheValTrp 304
Db 619 -----GTCTCACTCTGCGAGACT-----GACAGACTGTGATTTGA 581
Oy 305 AlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaTrpGluAlaSer 324
Db 580 GCTACATCTTACCGGAGGAGTCTGCT-----TCGCCAGTGGCCATACACAGCG 530
Oy 325 SerGlyGlyTyrLeuThrThrArgHisAsnGlyThrThrPheProAlaGlySerValSer 344
Db 529 TCAGCTGCC---AGCCCTCGG---AATGGAATGATCTCTCTCCAGGCGGCTG 479
Oy 345 TyrValLeuProGlnGlyPheAlaLeuGluArgTyrAspProAsnAspGlySerTrpThr 364
Db 478 TGGCGTTA-----CTGGCTCTGCGACCCAGGAGAC----- 446
Oy 365 AspPheAlaSerAlaGlyAspThrValThrPheArgGlnValAlaValAspGluValAl 384
Db 446 ----- 446
Oy 385 ValThrAsnAsnProAlaGlyGlySerAlaProThrPheThrValArgValProPro 404
Db 445 -----CGAGGCCAGATAGTGTGCCACCTTCATTA-----CCG 413
Oy 405 SerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeuGluThrArgProSerSer 424
Db 412 GACCTGTGTGGTGGCTTGAAGTGTCTTACAGCTACAGTGTGCGGCCCGCCCGCA--- 356
Oy 425 ArgArgGluGluLeuProMetProProAlaAspPheGlyGlnThrValAlaAsnAsnPro 444
Db 355 -----ATCCCTCCCGCGCCACACCTCCACCGGCTATCTTACACACA--- 311
Oy 445 LysIleGluGlnSerLeuLeuGlnThrLeuGlyCysTyrLeuValHisSerLysMet 464
Db 310 -----GCTCAGGAGGCTGTGTAACCTTACGTGACTGACTGCGCTTCA 269
Oy 465 ArgAsnProValPheGlnLeuThrProAlaSerSerPhe 477
Db 268 GATGAACCCCTCTTCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 230

```

RESULT 4

```

BM913440 1069 bp mRNA linear EST 12-MAR-2002
LOCUS AGENCOURT 6610241 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5475556
DEFINITION 5', mRNA sequence.
ACCESSION BM913440
VERSION BM913440.1 GI:19363819
KEYWORDS EST.

```

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 1069)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: DCTP/DTP
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LNCM1990 row: 3 column: 05
 High quality sequence stop: 734.
 Location/Qualifiers
 1. 1069
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5475556"
 /clone_lib="NIH_MGC_41"
 /issue_type="amelanotic melanoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH-MGC Library."
 BASE COUNT 214 a 321 c 293 g 223 t 18 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.859 Length: 1069
 Score: 111.50 Matches: 82
 Percent Similarity: 32.51% Conservative: 37
 Best Local Similarity: 22.40% Mismatches: 115
 Query Match: 3.30% Indels: 133
 DB: 14 Gaps: 17
 US-09-677-653a-50 (1-647) x BM913440 (1-1069)
 QY 86 Aspleuaspserserilegityttrpypheylstyreuasproalaglyalathr 105
 Db 87 AACCTGAGAACCGCTTACGCGGCTGTACGACCGACCTGAGCCGCGGCGCACGAG 146
 QY 106 Gluseralalargalavalglylutyrserysileproaspclyleuvalyspheser 125
 Db 147 GAGCGAAGCGCGCGC---GGCGAC-----GCCCTACGAGATGCTGCTATGAGTTTGAC 197
 QY 126 Valaspalagluileargluiletyrasnluclucysprovalvalthraspvalser 145
 Db 198 ATC-----TCCTTCACTTCAGTGCAG 218
 QY 146 Valproleuaspllyarglntrpserleuserilepheserphrometpheargthr 165
 Db 219 AAGAGACGATCCGACCCCTGACAGTG- 248
 QY 166 Alatyrralavalalaasnavalgluasnllygluwerterleuaspyalvalasnasp 185
 Db 249 -----CTAGATGCCATTGATCAG 266
 QY 186 leuilegluttrleuasnasleualaspttrpargttyrvalvalaspsergluintrip 205
 Db 267 ATG-----TGCTGCGCAGTGTGAGGACTTGGCGC----- 296
 QY 206 Ileasnphethrasnaspthrthrtyr----- 214

Db 297 -----CTCATGAGCGGCACTATGGGGGTCTAACCGGTCTCAATAAGCAGAAAGT 347
 QY 215 -----TyrValarglieargvalleuarargprothrtyrtyraspvalproasp 229
 Db 348 GCTGCAAGCATGGTGAAGGCCGAGGTGAGATGTGAGGCGGCTATGATGTCACCA 407
 QY 230 Prothrgluileuvalargthrvalseraspytyrargdeuthrtyrlysalaietthr 249
 Db 408 CCG----- 410
 QY 250 Cysglualaasnmetprothrleuvalaspglnglypheitpillegly----- 265
 Db 411 -----CCGATGGAGCCGACCACTTCCTTACAGCAACACTAAGAGAT 455
 QY 266 Glyglintyrala---leuthrprothrserleuproglntrtyraspvalserglualatyr 284
 Db 456 GCGAGGTATGCAAGCTCAGCAAGATCAGCTACCTCCTGTAGAG-TCTGAAAGATAC 514
 QY 285 AlaleuithrleuthrphthalargproserSerAlaAlaAlaAlaAlaAlaAlaAla 304
 Db 515 TATTCG-----CAGAGCTCTCCCTTCTGGAATGACAAAT 550
 QY 305 Alaglyleuproglnglytyrthralaproalaglythralatrpplugln----- 322
 Db 551 AGTTCCCAAGATCAGAGAGGGAAGCGTACTGATTCAGCCCAAGGCAAGCTCCG 610
 QY 323 -----Alasererlyglytyrleuthrtyrparghisnnglythrthrphepro 339
 Db 611 GGGCATTGTCACACATCTGGAGGGTCTCTGAAAGAGGCTATGATGAGCTGAA---CCT 667
 QY 340 AlaglyserValserTyrralValleuproglnlylphenalaleucluargrtyrasproasn 359
 Db 668 GCGGACTGCTATTTCCTTCTTATGATTTGACAAAGAACTTAAAGCTTATCAAGCCCAT 727
 QY 360 AspGlyserTrrpThrAspPheAlaSerAlaGlyAspThrValThrPheArglnValAla 379
 Db 728 GCAGTTTTCG-----GGGATGAAAGACAGCGGTGCCCAA---GCC 766
 QY 380 Valaspelualvalalvalthrasnasnproalaglygly----- 393
 Db 767 ATGGAAGCTGTGCTGCGCCAGGCAAGGCCANNAAGTGGAAGCGCGGAGAGATCTGT 826
 QY 394 -----SerAlaProThrPheThr-ValArg---ValproproserasnAl 407
 Db 827 CCCGAGAGCAACCCCTCCGCGCGGTTTGCCCTCCGCGCCCTCCACCTGGAGATGC 886
 QY 407 atyrrthrasnThrVal 412
 Db 887 CTTCCTGAACCCCATC 902
 RESULT 5
 LOCUS B0712831 1269 bp mRNA linear EST 16-JUL-2002
 DEFINITION AGENCOURT 8483154 NIH_MGC_129 Mus musculus cDNA clone IMAGE:5305997
 5', mRNA sequence.
 ACCESSION B0712831
 VERSION B0712831
 KEYWORDS B0712831.1 GI:21851730
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1269)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 CDNA Library Preparation: Resgen, Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LLM13721 row: c column: 22
High quality sequence stop: 264.

FEATURES

source

1. 1269

Location/Qualifiers

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:6305997"

/clone_11b="NIH_MGC_129"

/lab_host="DH10B (phage-resistant)"

/note="Organ: olfactory epithelium; Vector: PCMV-SPORT6.1.cdb; Site:1: Ecorey; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC library."

BASE COUNT 120 a 560 c 334 g 254 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 1.15 Length: 1269
Score: 111.50 Matches: 74
Percent Similarity: 38.27% Conservative: 32
Best Local Similarity: 26.71% Mismatches: 101
Query Match: 3.30% Indels: 70
DB: 14 Gaps: 15

US-09-677-653a-50 (1-647) x BQ712831 (1-1269)

OY 223 ProthrrTyraPvalProaspProthrgluLeuValArgThrValSerAspTyraArg 242
DB 15 CCTGACACAGCCATGTCATCTCTACGTCCTCCACCCGCGGCTTCC----- 65
OY 243 LeuthrTyraLalAlehrCysgluAlaAsnMetPro---ThrlauValaspGlnGly 261
DB 66 -----CCAGCTCCGAGCCCTCATCGCTGCGCATACGGG 101
OY 262 PheTrpLe-----GlyGlyGlnTyraLeuThrProThr-----SerLeu 275
DB 102 AGCGTGTACGGTCCCGGCTGGGAG---GCCCTACCCCGCATCTGTCGACGCTC 158
OY 276 ProGlnTyraPvalSerGluAlaThrValLeuThrPheAlaArgProSer 295
DB 159 CCCCCA-----GCAGCCGAGCTCATCTCCCTCCGCTGCGCGCCG 203
OY 296 SerAlaAlaLeuAlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAla 315
DB 204 TGGACAGGACCTGCGGCTGTGGGTGGGGGCGCGCTGACTCAGCTGTGT 263
OY 316 GlyThrProAlaTrpGluGlnAlaSerSerGlyGlyTyraLeuThrTrpArgHisAsnGly 335
DB 264 GGCTGACGAGCTGGGGGAGCCGGGGGAGCGGCGACCTGCTCGTCCATGACGGG 323
OY 336 ThrThrPheProAlaGlySerValSerTyraValLeuPro-GluGlyPheAlaLeuGluAr 355
DB 334 TCAGTCATGCTGACACGAGC---TCATACCCGCTGCTGCGGCGC----- 366
OY 355 gTyraAspPro-----AsnAspGlySerTrpThrAspPhe----- 366
DB 367 -----GACCTGCTGCTGCTGCGGATTTAAGTCTGCTGGTAGAGCCCGGCTGGCTTGC 422
OY 367 -----AlaSerAlaGlyAspThrValThrPheArgGlnAlaValAspGluValAla 385
DB 423 GGGCCCTGGGCGCTTGAACCCCGCGAGG-----CCCGGCTGCGCCCT 470
OY 385 lThrAsnAsnProAlaGlyGlySerAlaProThrPheThrValArgValProProSe 405
DB 471 CGACACCTACCCGCGGGGATGCCCCCTCTGCTTCCCTCGGCTCCCGCGG 530
OY 405 rAsnAlaTyThr-----AsnThrValPheArgAsnThrLeuLeuGlnThrAr 421
DB 531 AGCCGTTATTCGCCGCTGCCGACGACTACTACCCGCGACGCTGCTTGC----- 582

OY 421 gProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrValAl 441

DB 583 -----TTGCCATCCCGCTCCG----- 600

OY 441 aAsnAsnProLysIleGluGlnSerLeuLeuValThrLeuGlyCysTyraLeuValH 461

DB 601 -----TTTCGTGGGCGCGCGCGCGGATTTTTCATCTCCCGGCTGCTGCTCC 656

OY 461 sSerLysMetArgAsnProValPhe-----GlnLeuThrProAlaSer 475

DB 657 CACCCGATTCGTCCTCCCGCTGTGCGCGAGAGGTGTCGCCCGGAA 705

RESULT 6

BI082474 2128 bp mRNA linear EST 20-JUN-2001

LOCUS 60287815F1 NCI_CGAP_Man2 Mus musculus cDNA clone IMAGE:5009474 5',

DEFINITION mRNA sequence.

ACCESSION BI082474.1 GI:14500804

VERSION BI082474.1

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE NIH-MGC http://mgc.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM11055 row: n column: 03

High quality sequence stop: 42.

Location/Qualifiers

1. 2128

/organism="Mus musculus"

/strain="FVB/N-3"

/db_xref="taxon:10090"

/clone="IMAGE:5009474"

/clone_11b="NCI_CGAP_Man2"

/tissue_type="tumor, biopsy sample"

/dev_stage="5 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: PCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 576 a 872 c 330 g 350 t

ORIGIN

Alignment Scores:

Pred. No.: 4.58 Length: 2128
Score: 109.50 Matches: 74
Percent Similarity: 35.89% Conservative: 43
Best Local Similarity: 22.70% Mismatches: 133
Query Match: 3.25% Indels: 76
DB: 13 Gaps: 16

US-09-677-653a-50 (1-647) x BI082474 (1-2128)

OY 250 CysGluAlaAsnMetProThrLeuValaspGlnGlyPheTrpIleGlyGlnTyraAla 269

DB 352 TCCGCGCTTCATTCACACCACTTCCCTGCTGAGACACACGAGGAACCAATGGAAC 411

OY 270 Leu-----ThrProThrSerLeuProGlnTyraP-----ValSerGluAlaTy 284

DB 352 TCCGCGCTTCATTCACACCACTTCCCTGCTGAGACACACGAGGAACCAATGGAAC 411

```

Db 412 ATGATGGAACGTGAACCATTTTACCAACTACACGCGCGCTGTCACGAGGATTTAC 471
QY 285 -----AlaLeuH1sthrLeu-ThrPheAlaArgPr 294
Db 472 CACACGAAAGGGGGGACACACAGGGGGCCATTCACCTTTCATCACCCCCGGCC 531
QY 294 oSerAlaAlaAlaAlaAlaAlaPheValTrrPalaGlyLeuProGlnGlyThrAlaPr 314
Db 532 AGCTACTACAGCCCTCATC-----ATTCACCGCGCTGGTTCTTCAGCC 573
QY 314 oAlaGlyThrProAlaTrrPgluGlnAlaSerSerGlyGlyTyrLeuH1sthrParGlns-- 333
Db 574 CCCATCTCCCTCCCATATCAGCCCGCTAAGGACAGACTAATCATTCCTCCCTCAAC 633
QY 334 -----AsnGlyThrThrPheProAlaGlySerValSerTyrValLeuProGlnGlyPheAl 352
Db 634 ACCCAAGGGGAGTCATCAACCCACGCTCGACATCATTTGATCCCTTGGGCGACACT 693
QY 352 aLeuGluArgTyrAspProAsnAspGlySerTrrPThrAspPheAlaSerAlaGlyAspTh 372
Db 694 CCTTAAACACTTT-----AACGACGGCGCATAC 720
QY 372 rValThrPheArgGlnVal-----AlaValAspGluValValValThras 387
Db 721 ACTGGGAGGAGTCCGCTGACACACTAGAACGACACTTATACCGGCGCATACGCGATATCA 780
QY 387 nAspProAlaGlyGlyGly-----SerAlaProThrPheThrValArgValProProse 405
Db 781 TTACCTACTTAGGGCGGTATTCCTATCCCTCATTCGACCAAAATTTCTACTTCACCCCG 840
QY 405 r-----AsnAlaTyrThrAsnThrValPheArgAsnThrLeuGlu 419
Db 841 CCCAACAATAGTACCTTGGGGGACATCCTCCACATACAGTCCCTACACGACCAAGA 900
QY 419 uThrArgPro-----SerSerArgArgLeuGlnLeuProMetProProAlaAspPh 436
Db 901 GGCAAAACCCCAACCTTTCACCAACCATACCCCTTCATACCCGCCCTT----- 955
QY 436 eGlyGlnThrValAlaAsnAsnProLysIleGlnGlnSer-----LeuLeuLysGlnTh 454
Db 956 ----CACACACACCAACCACTTACACACATACACACATTCAGACACACTACTCAATTA 1011
QY 454 rLeuGlyCysTyrLeuValH1stSerLysMetArgAsnProValPheGlnLeuThrProAl 474
Db 1012 CCTAACCATATAC-----CACACCTCTAACCCCAACCTTCTACCCCTTACTCCACCC 1065
QY 474 aSerSerPheGlyAlaValSerPheAsnAsn-----Pr 485
Db 1066 ACATCTTTCMAAGCTTGGCAGCATTTTACTCCAGTCTTATCCGCGCTGACACCGTAC 1125
QY 485 oGlyTyrGluArgThrArgAspLeuProAspTyrThrGlyIleArgAspSerPheAspGl 505
Db 1126 AGCCGTACCGAGACAAACGATATCCGCCGACGCGACAGGTGTACAGGGGACCAAG 1185
QY 505 nAsnMetSerThrAlaValAlaH1stPheArgSerLeuSerH1stSerCysIleValTh 525
Db 1186 ATGACTAACCA-----CGCTCTCATCTCAATCCCATCC----- 1222
QY 525 rLysThrTyrGlnGly 530
Db 1223 ----ACCTACTACGCG 1234

```

RESULT 7
BC010575 3572 bp mRNA linear HTC 12-JUL-2001
LOCUS BC010575
DEFINITION Homo sapiens, similar to E74-like factor 1, clone IMAGE:4338775,
mRNA.
ACCESSION BC010575
VERSION BC010575.1 GI:14714847
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT Contact: MGC help desk
Email: ggaids-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalobedcm.tmc.edu.
Villalob, D.K., Luna, R.A., Hale, S.M., Hulik, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAC Plate: 12 Row: 0 Column: 13
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, similarity but not identity to protein

This clone has the following problem: incomplete processing.

```

FEATURES
source location/qualifiers
1..3572
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4338775"
/tissue_type="Lymph, Lymphoma"
/clone_id="NIH_MGC_85"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
BASE COUNT 1135 a 678 c 703 g 1056 t
ORIGIN
Alignment Scores:
Pred. NO.: 12.6 Length: 3572
Score: 109.00 Matches: 117
Percent Similarity: 35.22% Conservative: 82
Best Local Similarity: 20.71% Mismatches: 214
Query Match: 3.23% Indels: 153
DB: 11 Gaps: 22

```

US-09-677-653a-50 (1-647) x BC010575 (1-3572)

```

QY 97 LysTyrLeuAspProAlaGlyAlaThr-----GluSerAlaArgAlaValGlyGlu 113
Db 781 AAATATGACAGCTCACCGGAGCCTCATCACAGACAGCCTTAGAGGAAAAAGGAGA 840
QY 114 TyrSerLysIleProAspGlyLeuValLysPheSerValAspAlaGluIleArgGluIle 133
Db 841 AAAACTAACCA----- 855
QY 134 TyrAsnGluGluCysProValValThr-----AspValSerValPro-----LeuAsp 149
Db 856 ----CGACCAAGATTCGCCGACCTACGCGCAATATCTGTGAAGAAAGAAAACAAAGAT 912
QY 150 GlyArgGlnTrrPheSerLeuSerIlePheSerPheProMetPheArgThrAlaTyrValAla 169
Db 913 GGAAGGGAAGAACACATTTATCTTGGAGATT-----TTA 948
QY 170 ValAlaAsnValGluAsnLysGluMetSerLeuAspValValAsnAspLeuIleGluTrr 189
Db 949 CTGGCACTGCTCCAGCAAGAGCT-----ACTGTCTCTAATATCATCATCAAGTGG 996

```


Alignment Scores:

Pred. No.:	1 37	Length:	915
Score:	108.50	Matches:	80
Percent Similarity:	36.57%	Conservative:	33
Best Local Similarity:	25.89%	Mismatches:	115
Query Match:	3.22%	Indels:	81
	13	Gaps:	14

US-09-677-653a-50 (1-647) x B1413164 (1-915)

```

QY 228 PROAPProthrgluclyleuValArgThrValSerAspTyrGArgLeuThrTyr----- 245
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 CCGAGACAGCGTGGAGTACGTGAGAACCCCTGGACTCTCAGACCTGGAGCTTTATATGTC 178
QY 246 -----LysAlaIleThrgcsgluAlaasnMetPro 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 GGGGCCAGATGATTAAGAGATTAAAGAACACATACCTGCTCTGCAGCATCCCT 238
QY 256 ThrleuValasplnglphetrpIleGlyglntYrAlaLeuThrProthrsleu 275
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 TCGAGAACAGCGGCTCATCTACAGCCGGCTCTACAGACAGACAGACCTCC--- 295
QY 276 ProglntYrAspValsergluAlaTyrAlaLeuThrsleuThrPheAlaArgProser 295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 296 -----AGAACGCTGGG---GGAAAGGTATTCACCTGGTGGAGCGGCTCTCTCAG 346
QY 296 SerAlaAlaAlaLeuAlaPheValTrrAlaGlyLeuProglntglYrAlaProAla 315
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 347 ACTCAG-----CTTCCTCTGGAGCATCTCTGGGACA 379
QY 316 GlyThrProAlaTrrpGluglnAlaserSerGlyTyrTyrleuThrTrrpAghIsaNgly 335
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 GGGCTGGCTCA-----GCACTCATGTGGGGACCCCTG-----CCTGGC 421
QY 336 ThrThrPheProAlaGlySerVal-----SerTyrValLeuProglu 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 422 ACTCGGGGGCCCTGGGCTCTGTTCATGACCGGAATGCCAAGCATATGTCATGTTGGA 481
QY 350 GlyPheAlaLeuGlnArgTyrAspProAlasnAspGlySerTrrpThrAsp-PheAlaSerAl 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 482 ACCTTCAATCTT-----CCTAGTACGGGCTCTGCTGTTCATGTTCACTACATAC 529
QY 369 aglyAspThrValThrPheargGlnValAlaValaspluValValThr-AsnAsp 389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 530 ATGGAAACAGGCCCAATTCGA-----GTGAGCCCGGGTC 565
QY 389 roAlaGlyGlySerAla-----ProThrPheP 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 566 CGGCTGTGATGGCTCAGACATGATTAAGGATATACAGACTTACTGTCCCGGATGGAG 625
QY 399 hrValArgValProProSerAsnAlaTyrThrAsnThrValPheAlaGAsnThrleuG 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 626 TGTGAGGGGGGAACCCAGGCTCAGGCCACCCAGCCCGGAGAAC----- 675
QY 419 luthrArgProSerSerArgArgLeuGlnLeuProMetProProAlaAspPheGlyGlnT 439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 676 -----CGAGACTGTGGCTCAGAGACGGGGAAGCTTAACTACACAAA 718
QY 439 hrValAlaasnAsnProLysIleGlnSerleuLeuLysGlnThrleuGlyCysTyrL 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 719 ACATCAGAAACAGTCGAAAGTG---AAGCTCTCTCGAAGAGCCATGGGA----- 768
QY 459 euValIhtsSerLysMet-----ArgasnProValPheGlnLeuT 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 769 --GTCAAAGAAATGAGGAGGAACGTCGCCAACCCCAACCTCCAGAGACTTGGCGCCT 836
QY 472 hrProAlaSerSerPheGlyAlaValSerPheasnAsnProGlyTyrGluArgThrArga 492
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 827 CAAGGACACAGCTCCCGGACGGCCCGACAGCTCCCGGACGGCCAGGCTCCCGGACAGAC 886
QY 492 spleuProAspTrrpThrnglyle 499
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 887 AAGTGCGCCAGACAGCATT 909

RESULT 9
BM926649 1776 bp mRNA linear EST 12-MAR-2002
LOCUS
DEFINITION
AGENCOURT 6644658 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5767063
5', mRNA sequence.
ACCESSION
BM926649
VERSION
BM926649.1 GI:19377028
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 1776)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NIH-MGC http://mhc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cygabs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM12824 row: 1 column: 08
High quality sequence start: 70
High quality sequence stop: 413.

FEATURES

source

Location/Qualifiers
1..1776
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5767063"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."

BASE COUNT 455 a 524 c 462 g 325 t 10 others
ORIGIN

Alignment Scores:

Pred. No.:	4 29	Length:	1776
Score:	108.50	Matches:	77
Percent Similarity:	39.06%	Conservative:	23
Best Local Similarity:	30.08%	Mismatches:	106
Query Match:	3.22%	Indels:	50
	14	Gaps:	12

US-09-677-653a-50 (1-647) x BM926649 (1-1776)

```

QY 306 GlyLeuProGlnGlyGlyThr-AlaProAlaGlyThrProAlaTrrpGluglnAlaSerse 325
    ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 GGCATCTCTCTGTGAGACGGGCTCTCTCACACTCACCTCCCTTGGGGCATCTTC 124
QY 325 rglY---GlyTyrleuThrTrrpAghIsaSn-----GlyThrThrPheP 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 TGGAGAGGGGTGCTCACCACCATGATGGGGGATCCCGCCGTCGTCGGGGGCC 184
QY 339 calaGlySerVal-----SerTyrValLeuProGlnGlyPheAlaLe 353
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185 TGGGGCTCTGTTCATGACCGGAATGCCAACACTATGTATGTGTGAACCTCATCT 244
QY 353 uGluArgTyrAspProAsnAspGlySerTrrpThrAsp-PheAlaSerAlaGlyAspThrV 373

```



```

Db      822  TCCGGTGTCCGAGCGCGTGTCCGCTTCTCC 853
          |||||
          :
          |||||
RESULT 11
AL578091/c 869 bp  mRNA  linear  EST 16-FEB-2001
AL578091 LUTL_NFL006.PL2 Homo sapiens cDNA clone CS0DK003Y123 3
DEFINITION
prime, mRNA sequence.
ACCESSION
AL578091
VERSION
AL578091.1 GI:12941832
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 869)
AUTHORS
Li,W.B., Gruber,C., Jesse,J. and Polyes,D.
JOURNAL
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. 869
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK003Y123"
/tissue_lib="LUTL_NFL006.PL2"
/tissue_type="placenta"
/Note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 143 a 268 c 280 g 174 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 1.6 Length: 869
Score: 107.50 Matches: 59
Percent Similarity: 35.298 Conservative: 25
Best Local Similarity: 24.79% Mismatches: 90
Query Match: 3.19% Indels: 66
Gaps: 8
DB: 9
US-09-677-653a-50 (1-647) x AL578091 (1-869)
OY 260 GlnGlyPheTrpIleGlyGlnTyraLeuThProThrSerLeuProGlnTyraSP 279
||| |||||
584 CAGAACTTCTGATC-----AGCCTCCAGGAGCATG 552
OY 280 ValSerGlnAlaTyraLeuHisThrLeuThPheAlaArgProSerSerAlaAla 299
||| |||||
551 TGCAGTGAAGATGCGCTGACACATGCGATGATGACCGC-----510
OY 300 LeuAlaPheValTrpAlaGlyLeuProGlnGlyThraAlaPro--AlaGlyThPro 318
||| |||||
509 -----TGTGGAACGGGATGCCAGAGCGCGTACCTCCCGAGGTATGGGTACG 459
||| |||||
319 AlaTrpGlnAlaSerSerGlyGlyTyraLeuThTrpArgHisAsnGlyThThPhe 338
||| |||||
458 GCTGGCCACACAGATCAACACACCGAGGTGAGTGGACATCACACACCGGACATGA 399
||| |||||
OY 339 ProAlaGlySerValSerTyraValLeuProGlnGlyPheAlaLeuGlnAlaGly 358
||| |||||
398 CCATCCGCGAGCTGTCT-----GAACCGACATGACCT 366

```

```

OY 359 AsnAspGlySer-----TrpThrAspPhe 366
          :
          :
          :
Db 365 GAGAGGCGCCTTATGCTCTCTTCTTTCATCACCGCCCGACAGTGGAGGAGTC 306
OY 367 Alaser-----AlaGlyAspThrValThrPheArgGlnValAlaValAspGluVal 384
          :
          :
          :
Db 305 CCGGCTGTGTCAGTACAGTCCCATGCGTGTCTGTGACACCTTATTA-GATGTTTG 247
OY 385 ValThrAsnAsnProAlaGlyGlySerAlaProThrPheThrValAlaValProPro 404
          :
          :
          :
Db 246 GGATCAGAGACCCCGCAACACAGCAAGTCCACCCATTA-ATAACCCGACGACCGG 188
OY 405 SerAsnAlaTyraThrAsnThrValPheArgAsnThrLeuGlnLutThrArgProSerSer 424
          :
          :
          :
Db 187 TGGCGCTGGGACCTCTGACACAGTATG-----CCGGCGCC 152
OY 425 ArgArgLeuGlnLeuProMetProProAlaAspPheGlyGlnThrValAlaAsnAsnPro 444
          :
          :
          :
Db 151 AGGACAGCAGAACTCCCGCTGACACAGAGCGCTAGGGGTGGGCTCAGACCCACCT 92
OY 445 LysIleGlnInsLeuLeuLysGlnThrLeuGlnCysTyraLeuValHisSerLysMet 464
          :
          :
          :
Db 91 ACGGTC-----ATCTTGAAG 74
OY 465 ArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaValSerPhe 482
          :
          :
          :
Db 73 GGCAGCCCTGATGCTGCTACTGTGTCAGGCGAGTGGCAAGCTGCTGTCTTTC 20
RESULT 12
AG139016 686 bp DNA linear GSS 04-NOV-2001
LOCUS
pan troglodytes DNA, clone: PTB-153F24.R, genomic survey sequence.
ACCESSION
AG139016
VERSION
AG139016.1 GI:16668694
KEYWORDS
GSS.
SOURCE
Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
pan troglodytes
BAC library clone:PTB-153F24.R.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE
1
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE
BAC end sequences of library PTB
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 686)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE
Direct Submission
JOURNAL
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Sueni-ro-chou,Tsuri-mi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbpe@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rdb process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. 686
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-153F24.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 158 a 269 c 117 g 142 t
ORIGIN

```


